

RESEARCH ARTICLE

Aldehyde Dehydrogenase Gene Superfamily in *Populus*: Organization and Expression Divergence between Paralogous Gene Pairs

Feng-Xia Tian^{1,2}, Jian-Lei Zang^{1,2}, Tan Wang^{1,2}, Yu-Li Xie^{1,2}, Jin Zhang^{2,3*}, Jian-Jun Hu^{2*}

1 College of Life Science and Technology, Nanyang Normal University, Nanyang, Henan, 473061, China, **2** State Key Laboratory of Tree Genetics and Breeding, Key Laboratory of Tree Breeding and Cultivation of the State Forestry Administration, Research Institute of Forestry, Chinese Academy of Forestry, Beijing, 100091, China, **3** Research Institute of Wood Industry, Chinese Academy of Forestry, Beijing, 100091, China

* Zhang007jin@163.com (JZ); Hujj@caf.ac.cn (JJH)



OPEN ACCESS

Citation: Tian F-X, Zang J-L, Wang T, Xie Y-L, Zhang J, Hu J-J (2015) Aldehyde Dehydrogenase Gene Superfamily in *Populus*: Organization and Expression Divergence between Paralogous Gene Pairs. PLoS ONE 10(4): e0124669. doi:10.1371/journal.pone.0124669

Academic Editor: Arnar Pálsson, University of Iceland, ICELAND

Received: October 31, 2014

Accepted: March 16, 2015

Published: April 24, 2015

Copyright: © 2015 Tian et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was supported by National Nonprofit Institute Research Grant of CAF [CAFYBB2014ZX001-4] and [TGB2013009] to JJH and the China Postdoctoral Science Foundation [2014M550104] to JZ. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Abstract

Aldehyde dehydrogenases (ALDHs) constitute a superfamily of NAD(P)⁺-dependent enzymes that catalyze the irreversible oxidation of a wide range of reactive aldehydes to their corresponding nontoxic carboxylic acids. ALDHs have been studied in many organisms from bacteria to mammals; however, no systematic analyses incorporating genome organization, gene structure, expression profiles, and *cis*-acting elements have been conducted in the model tree species *Populus trichocarpa* thus far. In this study, a comprehensive analysis of the *Populus* ALDH gene superfamily was performed. A total of 26 *Populus* ALDH genes were found to be distributed across 12 chromosomes. Genomic organization analysis indicated that purifying selection may have played a pivotal role in the retention and maintenance of *PtALDH* gene families. The exon-intron organizations of *PtALDHs* were highly conserved within the same family, suggesting that the members of the same family also may have conserved functionalities. Microarray data and qRT-PCR analysis indicated that most *PtALDHs* had distinct tissue-specific expression patterns. The specificity of *cis*-acting elements in the promoter regions of the *PtALDHs* and the divergence of expression patterns between nine paralogous *PtALDH* gene pairs suggested that gene duplications may have freed the duplicate genes from the functional constraints. The expression levels of some ALDHs were up- or down-regulated by various abiotic stresses, implying that the products of these genes may be involved in the adaptation of *Populus* to abiotic stresses. Overall, the data obtained from our investigation contribute to a better understanding of the complexity of the *Populus* ALDH gene superfamily and provide insights into the function and evolution of ALDH gene families in vascular plants.

Introduction

Endogenous aldehydes are intermediates or byproducts in a range of fundamental biochemical pathways that are generated during the metabolism of carbohydrates, vitamins, steroids, amino acids, and lipids [1, 2]. When produced in excessive amounts, these aldehydes can have detrimental effects on cellular metabolism because of their chemical reactivity [3–5]. Therefore, cellular levels of aldehydes need to be regulated to maintain normal developmental processes. One of the major detoxification pathways of aldehyde molecules involves the oxidation of the carbonyl groups to carboxylic acids by NAD(P)⁺-dependent enzymes aldehyde dehydrogenases (ALDHs; enzyme class EC: 1.2.1.3) [1–3, 6].

ALDHs are found in both prokaryotes and eukaryotes [2, 3, 6–8]. The ALDH Gene Nomenclature Committee has established specific criteria for cataloguing deduced ALDH protein sequences [6]; namely, protein sequences that share more than 40% identity with other ALDH sequences compose a family, and sequences that share more than 60% identity compose a sub-family. ALDH protein sequences that share less than 40% identity would form a new family. Previous classifications of the *ALDH* gene superfamily in eukaryotes have identified 24 protein families based on sequence identity [7–10]. Among the 24 ALDH families, 14 (ALDH2, ALDH3, ALDH5, ALDH6, ALDH7, ALDH10, ALDH11, ALDH12, ALDH18, ALDH19, ALDH21, ALDH22, ALDH23 and ALDH24) contain members from plant species and seven (ALDH11, ALDH12, ALDH19, ALDH21, ALDH22, ALDH23 and ALDH24) are unique to plants [10].

Of the plant *ALDHs* that have been characterized to date, the majority have been implicated in diverse pathways and appear to play crucial roles in plant growth and development. For example, the maize *ALDH2* gene *rf2* is required for male fertility [11] and the rice *ALDH7* is essential for seed maturation and viability [12]. Many of the plant *ALDHs* are responsive to various environmental stresses, including dehydration, high salinity, heat, water logging, oxidative stress, and heavy metals, suggesting possible roles for these genes in improving stress tolerance [13–15]. Several studies have found that overexpression of some plant *ALDHs* enhanced plant tolerance to diverse types of abiotic and biotic stresses [13–17]. Most of these studies have been performed in model species such as *Arabidopsis* [8] and rice [18], and, until now, little attention has been paid to woody species like *Populus*.

During plant evolution, gene families have undergone copy number selection via duplications, transpositions and/or deletions [19]. Gene duplication and subsequent gene retention or loss (fractionation) are often attributed to recent and/or ancient whole genome polyploidy events, for example, at the origin of seed plants and angiosperms [20]. Whole-genome duplications can buffer gene functions by increasing genetic redundancy and hence contribute to sub- or neo-functionalization, which can drive genetic innovation [21]. For example, paralogous genes derived from a whole-genome duplications that encode structurally similar enzymes have been shown to evolve towards extended substrate specificities or to catalyze novel reactions, whereas the ancestral gene retains its original function [22].

Populus species are perennial trees that frequently undergo seasonal variations and various environmental stresses. The completion of the *Populus trichocarpa* genome sequencing project in 2006 made *P. trichocarpa* a suitable model for woody plants [23]. An analysis of the *P. trichocarpa* genome found that a whole-genome duplication event may have occurred recently (in evolutionary terms) in the stem lineage of the Salicaceae family, about 60 to 65 million years ago, in addition to another, much more ancient large-scale duplication event shared by *Populus* and *Arabidopsis* [23]. The complex history of genome duplications and chromosomal rearrangements in *Populus* provide an opportunity to study gene family expansion patterns over the course of genome evolution [24]. In a previous study, Brocker et al. [25] identified 26

Populus ALDHs based on the *P. trichocarpa* genome V2.2; however, no systematic analyses of the evolution and expression patterns were analyzed. To determine the structure-function relationship of the ALDHs in *Populus*, we performed detailed systematic analyses of the genome organization, gene structure, expression compendium, and *cis*-acting elements in *P. trichocarpa*. In this study, we report comprehensive phylogenetic and evolutionary analyses of the 26 members of the ALDH superfamily in *Populus*, including their expression profiles in different tissues and their responses under various abiotic stresses. The specificity of *cis*-acting elements between paralogous *Populus* ALDH gene pairs were applied to investigate the divergence of their expression patterns to help understand how paralogous genes play different roles in various biological processes and stress responses. Our results may provide the insights to further investigate the functions of the ALDHs in *Populus* species.

Materials and Methods

Characteristics of *Populus* ALDH genes

A previous study identified 26 *Populus* ALDH genes based on the *P. trichocarpa* genome V2.2 [25]. We also checked the latest *P. trichocarpa* genome V3.0 by BLASTP and no more ALDH members were identified. Total of 26 *Populus* ALDHs were named according to Brocker et al. and their coding sequences were download from *P. trichocarpa* genome V3.0 (http://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Ptrichocarpa). To characterize the members of *Populus* ALDH superfamily, WoLF PSORT (<http://wolfsort.org>) was used to predict protein subcellular localization [26]. The pI and molecular weight were estimated using the Compute pI/Mw tool from ExpASy (http://web.expasy.org/compute_pi).

Sequence alignments and phylogenetic analyses

Multiple alignment of ALDH protein sequences from *P. trichocarpa* and *A. thaliana* were performed using the Clustal X2.1 program [27]. The phylogenetic trees were constructed using the neighbor-joining method [28] in the MEGA package V5.2 [29] with bootstrap values from 1,000 replicate indicated at each node. The full length sequences of ALDH proteins used for phylogenetic analysis were listed in [S1 Table](#).

Bioinformatics analysis of *Populus* ALDH genes

The exon and intron structures were illustrated using Gene Structure Display Server (GSDS, <http://gsds.cbi.pku.edu.cn>) [30] by aligning the cDNA sequences with the corresponding genomic DNA sequences from Phytozome (<http://phytozome.jgi.doe.gov/pz/portal.html#>). The chromosomal locations of the ALDH genes were determined using the *Populus* genome browser (http://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Ptrichocarpa). Tandem duplicated *PtALDH* genes were defined as adjacent homologous ALDH genes on the *Populus* chromosomes, with no more than one intervening gene. For synteny analysis, synteny blocks within the *Populus* genome and between *Populus* and *Arabidopsis* genomes were downloaded from the Plant Genome Duplication Database (PGDD, <http://chibba.agtec.uga.edu/duplication/>) [31] and those containing *Populus* ALDH genes were identified. The chromosomal locations of ALDH genes were drawn using Circos software [32]. To analyze the putative *cis*-acting regulatory elements, -1,000 nt of the upstream to +200 nt of the downstream of transcription start site (TSS) were searched using PlantCARE database [25].

Publicly available microarray data analyses

The microarray data for various tissues and developmental stages available at NCBI Gene Expression Omnibus (GEO) database [33] under the series accession numbers GSE13990 and GSE13043 were used for the tissue-specific expression analysis. The series GSE13990 includes Affymetrix microarray data from nine different tissue samples representing three biological replicates [34], whereas series GSE13043 contains NimbleGen microarray data from five stem internodes (IN) from the apical bud to the base of the shoot (IN2 to IN5, IN9) in two biological replicates [35]. For abiotic and hormonal treatments, Affymetrix microarray data available in the NCBI GEO database under the series accession numbers GSE13109 (hypoxia), GSE17225 (drought), GSE26199 (heat) and GSE16786 were analyzed [36, 37]. GSE16786 is composed of the following five subsets: GSE14893 (nitrogen limitation, genotype 1979), GSE14515 (nitrogen limitation, genotype 3200), GSE16783 (1 week after leaf wounding), GSE16785 (90 h after leaf wounding), and GSE16773 (methyl jasmonate-elicited suspension cell cultures). Probe sets corresponding to *Populus ALDH* genes were identified using the online Probe Match tool POParray (<http://aspendb.uga.edu/poparray>). The probe sets corresponding to *Populus ALDH* genes were listed in S2 Table.

In Affymetrix GeneChip array, oligonucleotides of length 25 bp are used to probe genes. Typically, a gene is represented by a probe set composed of 11–20 probe pairs of these oligonucleotides [38]. To check the specificity of *PtALDHs* probe sets, we compared the recognition sites of *PtALDHs* probe sets between the paralogous genes. Except the probe sets corresponding to *PtALDH2B4* have four nonspecific binding sites in *PtALDH2B6*, all the other probe sets are highly specific to their corresponding genes (S1 Fig). So these probe sets could be used to reflect the real expression pattern.

Plant material, RNA isolation and Real-time qRT-PCR

1-year-old *P. trichocarpa* grown in a growth chamber under long-day conditions (16 h light/8 h dark) at 23–25°C. Plant materials for qRT-PCR in different tissues (YL—young leaf, ML—mature leaf, PS—primary stem, SS—secondary stem, and R—root) were collected from 84K. Samples were frozen immediately in liquid nitrogen, and stored at -80°C for further analysis. Three biological replicates were performed.

Total RNA was extracted using the RNeasy Plant Mini Kit (Qiagen) with on-column treatment with RNase-free DNase I (Qiagen) to remove any contamination of genomic DNA. First-strand cDNA synthesis was carried out with approximately 1 µg RNA using the SuperScript III reverse transcription kit (Invitrogen) and random primers according to the manufacturer's procedure. Primers with melting temperatures of 58–60°C and amplicon lengths of 100–250 bp were designed using Primer3 software (<http://frodo.wi.mit.edu/primer3/input.htm>). All primer sequences are listed in S3 Table.

Real-time qRT-PCR was conducted on 7500 Real Time PCR System (Applied Biosystems, CA, USA) using SYBR Premix Ex Taq Kit (TaKaRa, Dalian, China) according to the manufacturer's instructions. Reactions were prepared in a total volume of 20 µl containing: 10 µl of 2×SYBR Premix, 2 µl of cDNA template, 0.4 µl of each specific primer to a final concentration of 200 nM. The reactions were performed as the following conditions: initial denaturation step of 95°C for 30 s followed by two-step thermal cycling profile of denaturation at 95°C for 10 s, and combined primer annealing/extension at 60°C for 34 s for 40 cycles. To verify the specificity of each primer pair, a melting curve analysis was performed ranging from 60°C to 95°C with temperature increasing steps of 0.06°C/s (5 acquisitions per °C) at the end of each run. The final threshold cycle (Ct) values were the mean of eight values including two biological

replicates for each treatment and four technical replicates. The *PtActin* and *PtTubulin* gene were used as internal controls.

Statistical analysis

The statistical significance of differences in measured parameters was tested by using the procedures of DPS (Zhejiang University, China). Differences between the means among different tissues or gene pairs were compared using Duncan test and Fisher's protected least significant difference (LSD) test at 0.05 probability levels.

Results and Discussion

Characteristics of *ALDH* gene families in *Populus*

Previous study identified 26 *ALDH*s in *P. trichocarpa* [25]. The *Populus* *ALDH* superfamily is larger than the *Arabidopsis* *ALDH* superfamily, the number of *ALDH*s in the *P. trichocarpa* genome was found to be consistent with 1.4–1.6 putative *Populus* homologs for each *Arabidopsis* gene according to previous comparative genomics studies [23]. The *ALDH*s identified in *P. trichocarpa* encode proteins that range from 401 to 958 amino acids (aa) in length, with predicted isoelectric points (pIs) from 5.44 to 9.12 (Table 1). In this study, we named the 26 *Populus* *ALDH* genes according to Brocker et al. [25]. The *ALDH* proteins from *Populus* were grouped into 10 families based on their protein sequence identities and phylogenetic relationships with *Arabidopsis* *ALDH*s (Fig 1 and S4 Table). Seven of the 10 *ALDH* families in *Populus* were represented by more than one gene (*ALDH3*, six genes; *ALDH2*, four genes; *ALDH6*, four genes; *ALDH11*, three genes; *ALDH7*, *ALDH10*, and *ALDH18*, two genes), whereas the remaining three families (*ALDH5*, *ALDH12*, and *ALDH22*) each were encoded by a single-copy gene.

To investigate the evolutionary relationships of *ALDH* proteins from different organisms, we summarized numbers of gene family members for each individual *ALDH* family in *P. trichocarpa* and seven other plant species (*Arabidopsis thaliana* [8], *Vitis vinifera* [10], *Zea mays* [39], *Oryza sativa* [18], *Physcomitrella patens*, *Chlamydomonas reinhardtii*, and *Ostreococcus tauri* [40]), three mammals (*Homo sapiens*, *Mus musculus*, and *Rattus norvegicus*) and fungi [7] (S5 Table). Plant *ALDH*s are grouped into 13 families: *ALDH2*, *ALDH3*, *ALDH5*, *ALDH6*, *ALDH7*, *ALDH10*, *ALDH11*, *ALDH12*, *ALDH18*, *ALDH21*, *ALDH22*, *ALDH23*, and *ALDH24*. *Populus* and other vascular plants share 10 common core *ALDH* families (i.e. families 2, 3, 5, 6, 7, 10, 11, 12, 18, and 22), suggesting that these 10 families may have evolved before the divergence of monocots and eudicots. Moreover, eight of the 10 core families also are shared by terrestrial plants and algae (families 2, 3, 5, 6, 10, 11, and 12), suggesting that these families have ancient origins that predate the transition of aquatic plants onto land. Interestingly, it was reported previously that *ALDH19* was also unique to plants; however, no *Populus* homologue of *ALDH19* was detected in either this study or other studies of vascular plants. To date, the only known member of *ALDH19* family is a single gene from tomato, suggesting that *ALDH19* may have evolved specifically in this lineage [41].

It is worth noting that vascular plants such as *P. trichocarpa*, *Z. mays*, and *V. vinifera* have more *ALDH*s than animals and fungi. To date, *Populus* *ALDH* families are the most expanded with 26 genes compared with the *ALDH* families in other well characterized plants (24 *ALDH*s in *Z. mays*, 23 in *V. vinifera*, 21 in *O. sativa*, 16 in *A. thaliana*, 20 in *P. patens*, eight in *C. reinhardtii*, and six in *O. tauri*). Unlike animals, plants cannot move to avoid exposure to environmental stresses and, as a result, plants may require many stress-response proteins to protect them when exposed to abiotic and biotic stresses [10, 42]. Compared with other annual plant species, *Populus* species undergo secondary growth, seasonal variation, and are exposed frequently to various environmental stresses. The expanded *ALDH* families in *Populus* imply that

Table 1. Populus ALDH genes and superfamilies.

Family	Gene Name	Gene Locus	CDS (bp)	ORF (aa)	pI	MW (kDa)	SubLoc*
Family 2	<i>PtALDH2B4</i>	Potri.012G078700.1	1611	536	8.74	58.65	mito: 8.5, chlo_mito: 7.5, chlo: 5.5
	<i>PtALDH2B6</i>	Potri.015G074100.1	1575	524	8.68	56.89	mito: 9.5, chlo_mito: 7.5, chlo: 4.5
	<i>PtALDH2B7</i>	Potri.002G189900.1	1623	540	6.11	58.84	mito: 9.5, chlo_mito: 6.5, chlo: 2.5
	<i>PtALDH2C4</i>	Potri.018G075000.1	1245	414	6.76	45.10	cysk: 8.0, cyto: 5.0
Family 3	<i>PtALDH3F1</i>	Potri.007G017100.1	1443	480	6.97	53.62	cyto: 5.0, golg: 5.0, plas: 2.0, chlo: 1.0
	<i>PtALDH3H1</i>	Potri.005G179300.1	1467	488	7.08	53.50	cyto: 9.0, chlo: 2.0, cysk: 2.0
	<i>PtALDH3H4</i>	Potri.002G081800.1	1467	488	7.08	53.97	cyto: 8.0, cysk: 3.0, chlo: 2.0
	<i>PtALDH3H5</i>	Potri.001G412900.1	1647	548	8.19	60.68	nucl: 6.0, chlo: 4.0, cyto: 3.0
	<i>PtALDH3H6</i>	Potri.005G069800.1	1467	488	5.68	54.12	cysk: 8.0, cyto: 5.0
	<i>PtALDH3J1</i>	Potri.001G259100.1	1206	401	9.12	44.21	chlo: 8.0, cyto: 3.0, extr: 2.0
Family 5	<i>PtALDH5F1</i>	Potri.010G174000.1	1611	536	8.46	57.50	mito: 10.0, chlo: 4.0
Family 6	<i>PtALDH6B3</i>	Potri.009G078600.1	1629	542	6.68	58.01	mito: 10.5, chlo_mito: 7.5, chlo: 3.5
	<i>PtALDH6B4</i>	Potri.001G283100.1	1629	542	7.00	57.90	mito: 12.0, chlo: 2.0
	<i>PtALDH6B5</i>	Potri.009G078700.1	2142	713	8.36	77.64	nucl: 7.0, chlo: 2.0, cyto: 2.0, cysk: 2.0
	<i>PtALDH6B7</i>	Potri.005G147700.1	2877	958	6.76	104.53	nucl: 11.0, chlo: 1.0, vacu: 1.0
Family 7	<i>PtALDH7B4</i>	Potri.003G067700.1	1527	508	5.71	54.69	nucl: 7.0, cyto: 3.0, chlo: 1.0, plas: 1.0, pero: 1.0
	<i>PtALDH7B5</i>	Potri.001G167100.1	1527	508	5.55	54.89	nucl: 7.0, chlo: 3.0, cyto: 2.0, plas: 2.0
Family 10	<i>PtALDH10A8</i>	Potri.012G075600.1	1512	503	5.46	54.93	chlo: 8.0, cyto: 4.0, pero: 2.0
	<i>PtALDH10A9</i>	Potri.015G070600.1	1620	539	5.44	59.13	chlo: 7.0, cyto: 4.0, pero: 2.0
Family 11	<i>PtALDH11A3</i>	Potri.018G109700.1	1521	506	7.86	54.30	cysk: 8.0, cyto: 5.0
	<i>PtALDH11A4</i>	Potri.006G186800.1	1497	498	7.84	53.64	cysk: 10.0, cyto: 4.0
	<i>PtALDH11A5</i>	Potri.018G020600.1	1497	498	6.72	53.44	cysk: 9.0, cyto: 5.0
Family 12	<i>PtALDH12A1</i>	Potri.015G064200.1	1704	567	7.65	62.97	mito: 8.5, chlo: 5.0, cyto_mito: 5.0
Family 18	<i>PtALDH18B1</i>	Potri.010G198400.1	2160	719	6.46	77.78	E.R.: 5.0, chlo: 3.0, nucl: 1.0, cyto: 1.0, mito: 1.0, plas: 1.0, extr: 1.0
	<i>PtALDH18B2</i>	Potri.008G060200.1	2148	715	6.05	77.46	chlo: 4.0, E.R.: 4.0, nucl: 2.0, cyto: 1.0, plas: 1.0, extr: 1.0
Family 22	<i>PtALDH22A1</i>	Potri.008G106000.1	1785	594	6.84	65.70	chlo: 4.0, cyto: 3.5, cyto_nucl: 2.5, plas: 2.0, vacu: 2.0, mito: 1.0

Notes: Gene loci are obtained from the Phytozome website (<http://www.phytozome.net>).

* PSORT predictions: plas (plasma membrane), vacu (vacuolar membrane), cyto (cytosol), chlo (chloroplast), cysk (cytoskeleton), pero (peroxisome), nucl (nuclear), E.R. (endoplasmic reticulum), mito (mitochondrion), golg (golgi), extr (extracellular).

The numbers indicate the number of nearest neighbors to the query which localize to each site.

doi:10.1371/journal.pone.0124669.t001

the *ALDHs* may be involved in developmental processes or improving stress tolerance in woody species.

Phylogenetic and gene structure analyses of *Populus ALDH* genes

To examine the phylogenetic relationships among the *Populus ALDH* genes and other plant species, we generated a phylogenetic tree by aligning the full-length ALDH protein sequences identified in *P. trichocarpa* and *A. thaliana* [8]. As shown in Fig 1A, the ALDHs from the same families tended to cluster together. Among these families, *ALDH18* was the most distantly related family in the phylogeny. This finding is consistent with previous research in rice, which indicated that two *OsALDH18* proteins had the greatest degree of sequence divergence from

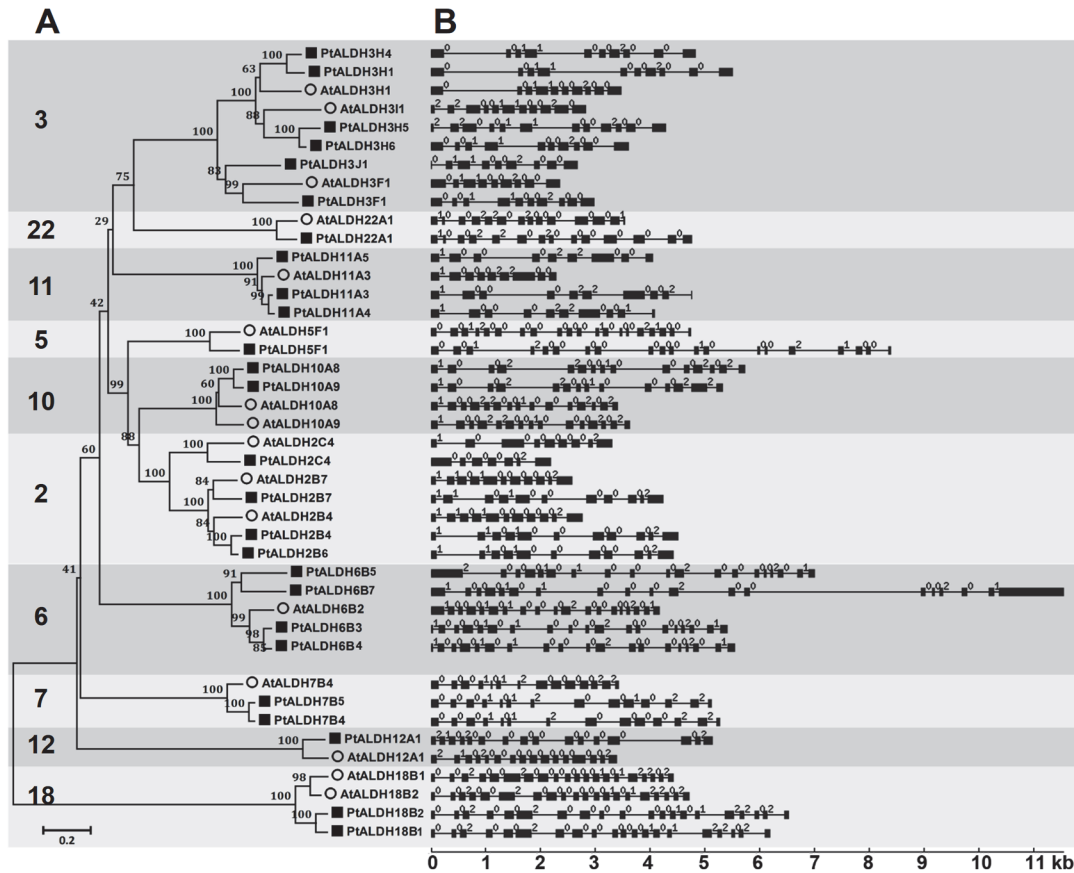


Fig 1. Phylogenetic analysis (A) and exon-intron structures (B) of *Arabidopsis* and *Populus* ALDH genes. (A) Multiple alignment of ALDH proteins from *A. thaliana* and *P. trichocarpa* was performed using Clustal X2.1. Phylogenetic tree was constructed using full-length protein sequences by the neighbor-joining (NJ) method with 1,000 bootstrap replicates using MEGA 5.2. Numbers above branches of the tree indicate bootstrap values. (B) Exon-intron structures of the ALDH genes. Only coding exons, represented by black boxes were drawn to scale. Lines connecting two exons represent introns. The numbers indicate the splicing phases of the ALDHs: 0, phase 0; 1, phase 1; and 2, phase 2.

doi:10.1371/journal.pone.0124669.g001

the other ALDH families and did not contain the conserved ALDH active sites [18]. A likely reason for this observation is that members of the ALDH18 family may be involved in a variety of biological processes, which require that a very diverse range of substrates can be recognized in a sequence- and/or structure-specific manner.

Furthermore, we analyzed the exon-intron organization in the 26 *Populus* ALDHs and 16 *Arabidopsis* ALDHs to investigate their structural diversity (Fig 1B). Exon-intron structural divergence within families plays a pivotal role in the evolution of multiple gene families. Generally, the positions of some spliceosomal introns were conserved in orthologous genes and, in many cases, conservation of the exon-intron organization in paralogous genes was high enough to reveal the evolutionary relationship between introns [43]. As shown in Fig 1B, most of the members in some of the families (2, 3, 7, 10, 11, and 18) had the same number of exons and nearly identical exon lengths. We also examined intron phases with respect to codons and found that the intron phases were remarkably well conserved among family members, whereas the intron arrangements and intron phases were distinct between families (Fig 1B). The high degree of sequence identity and similar exon-intron structures of ALDHs within each family suggested that *Populus* ALDH families may have undergone gene duplications throughout evolution, resulting in ALDH gene families that contain multiple copies of similar genes with

functions that partially or completely overlap. Previous studies reported that *ALDHs* from rice, grape, and *Arabidopsis* had highly conserved exon-intron structures [10, 18]. Here, we compared the exon-intron structures of the *Populus* and *Arabidopsis* *ALDHs* and found that the gene structures were conserved not only within a species but also across these two species (Fig 1B). However, we also identified exons that had been gained or lost during the evolution of several of the *ALDHs*. One such example is the *ALDH3* gene family in which *PtALDH3H5* and *PtALDH3H6* seem to have acquired one additional exon at their 3'-end, while *PtALDH3H6* lost the first two exons and *PtALDH2C4* may have lost two exons at their 5'-ends.

Chromosomal location and expansion patterns of *Populus* *ALDH* genes

In silico mapping of the gene loci showed that the 26 *Populus* *ALDHs* were mapped unevenly to 12 of 19 *Populus* chromosomes (chr). Chr1 had the largest number of four *ALDHs* followed by three *ALDHs* on chr5, chr15, and chr18. In contrast, only one *ALDH* mapped to chr3, chr6, and chr7 and two *ALDHs* mapped to chr2, chr8, chr9, chr10, and chr12 (Fig 2). No substantial clustering of the *Populus* *ALDH* genes was observed, even on the chromosomes with high densities of *ALDHs*.

Previous analysis indicated that the *Populus* genome may have undergone at least three rounds of genome-wide duplications followed by multiple segmental duplication, tandem duplication, and transposition events such as retroposition and replicative transposition [44]. To determine the evolutionary relationships among the *ALDHs*, we mapped *Populus* *ALDHs* to the duplicated blocks reported previously [23]. The distributions of the *ALDHs* relative on the corresponding duplicated blocks are shown in Fig 2. Within the duplicated blocks that were reported to be associated with recent salicoid duplication events, 92.3% (24 of 26) of the *Populus* *ALDHs* were preferentially retained duplicate genes that were located in both duplicated regions; only two *ALDHs* (*PtALDH3F1* and *3H5*) were located outside any of the duplicated blocks. Eight duplicated blocks contained *ALDHs* (*PtALDH2B7*, *2C4*, *3J1*, *3H6*, *5F1*, *6B7*, *12A1*, and *22A1*) on only one of the blocks and lacked duplicate genes on the corresponding blocks. These results indicated that dynamic rearrangement may have occurred following the segmental duplication and this resulted in the loss of some genes.

Two tandem *ALDH* gene duplications have been reported in rice (*OsALDH2-1/2-2* and *OsALDH3-1/3-2*) [18] and grape (*VvALDH5F1/5F2/5F3* and *VvALDH6B3/6B5*) [10]. In the present study, we also identified tandem duplications in the *Populus* *ALDH6* gene family (*PtALDH6B3/6B5*) (Fig 2 and Table 2). Analysis of *ALDH* paralogous pairs showed that seven of 10 gene pairs were located in conserved positions on segmental duplicated blocks, indicating that these genes might have been generated by genome duplication (Fig 2 and Table 2). The high retention rate (14/26, 53.8%) of the duplicated *ALDHs* was consistent with recent reports of other gene families in *Populus* [45–47]. In summary, the seven *Populus* multi-member *ALDH* families (Table 2) all were associated with either segmental or tandem duplication events, indicating that segmental and tandem duplications may have played important roles in the expansion of *ALDHs* in *Populus*.

Duplication and evolution analysis of the *Populus* *ALDH* genes

Duplicated genes may undergo divergent fates such as nonfunctionalization (loss of original functions), neofunctionalization (acquisition of novel functions), or subfunctionalization (partition of original functions) [48, 49]. To determine whether positive selection was involved in the divergence of *ALDHs* after duplication, the nonsynonymous (K_a) to synonymous (K_s) ratios were calculated for paralogous *PtALDH* gene pairs [50]. $K_a/K_s = 1$ indicates neutral selection, $K_a/K_s > 1$ indicates accelerated evolution with positive selection, and $K_a/K_s < 1$ indicates

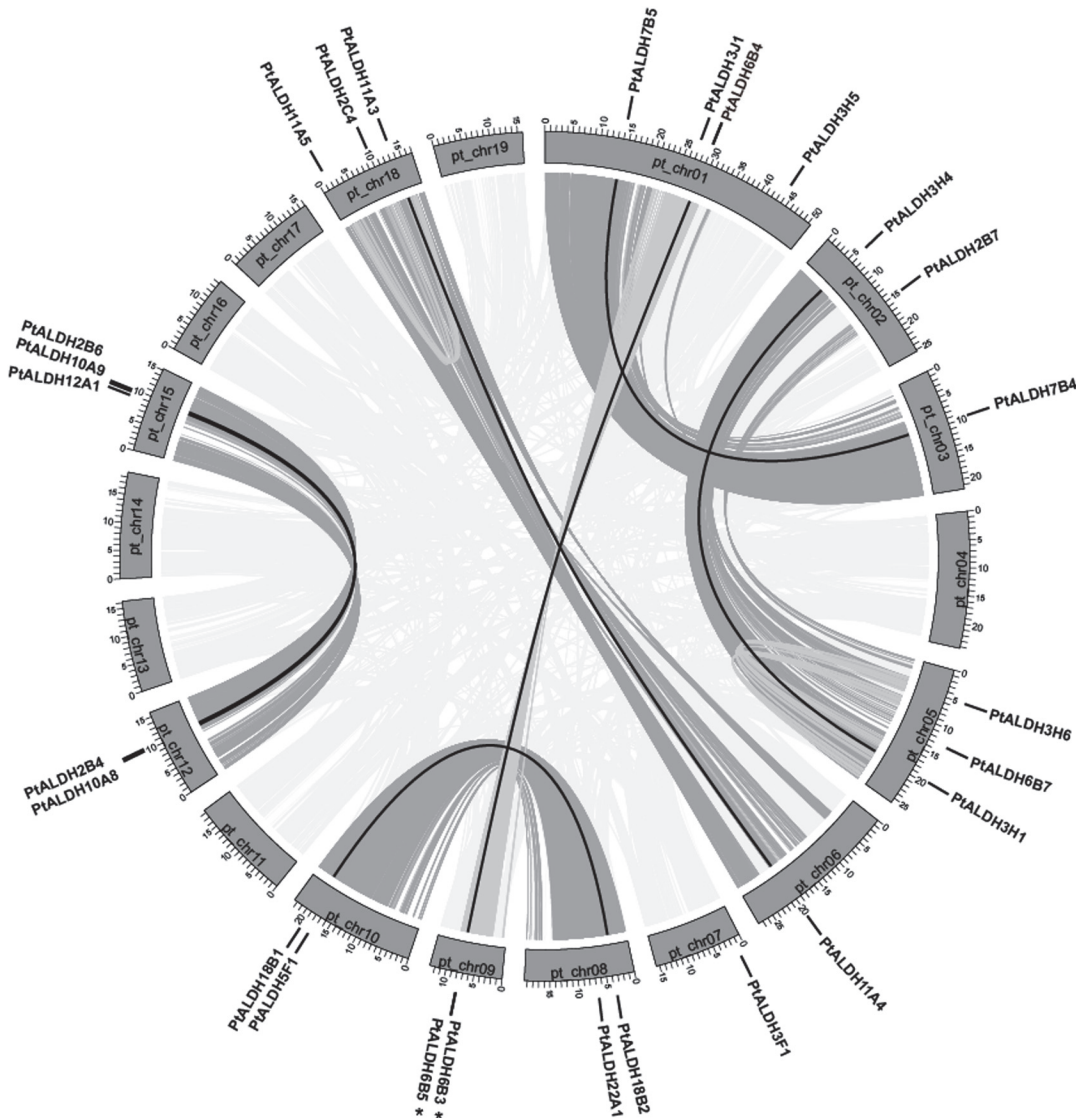


Fig 2. Distribution and synteny of ALDH genes on *Populus* chromosomes. Nineteen *Populus* chromosomes (chr01-chr19) are depicted as grey bars. Chromosome numbers are indicated in centers of each bar. *Populus* ALDH genes are indicated by vertical black lines outer the circles. Duplicate pairs formed by whole or segmental genome duplication are connected by black lines. Grey lines connected between chromosomes denote syntenic regions of the *Populus* genome.

doi:10.1371/journal.pone.0124669.g002

purifying selection [51]. The K_a/K_s ratios of all 10 *PtALDH* gene pairs were less than 1 (Table 2), implying that *Populus* ALDH gene pairs may have evolved mainly under the influence of purifying selection.

Based on the divergence rate of 9.1×10^{-9} synonymous mutations per synonymous site per year proposed previously for *Populus* [52], we estimated the evolutionary dates of the segmental duplication events using K_s as the proxy for time (Table 2). We found that seven of the 10 paralogous pairs (*PtALDH2B4/2B6*, *PtALDH3H1/3H4*, *PtALDH3H51/3H6*, *PtALDH6B3/6B4*, *PtALDH7B4/7B5*, *PtALDH11A3/11A4*, *PtALDH18B1/18B2*) had very consistent K_s values (from 0.188 to 0.281), suggesting that these duplication events occurred in *Populus* within the last 10.35 to 15.44 million years. This period is consistent with the time (13 million years)

Table 2. Divergence between paralogous ALDH genes pairs in *Populus*.

Family	Gene 1	Gene 2	Duplication	Ka	Ks	Ka/Ks	Date (million years ago)
2	<i>PtALDH2B4</i>	<i>PtALDH2B6</i>	W	0.047	0.281	0.169	15.44
3	<i>PtALDH3H1</i>	<i>PtALDH3H4</i>	W	0.065	0.206	0.316	11.34
3	<i>PtALDH3H5</i>	<i>PtALDH3H6</i>	O	0.053	0.224	0.235	12.32
6	<i>PtALDH6B3</i>	<i>PtALDH6B4</i>	W	0.033	0.218	0.151	11.99
6	<i>PtALDH6B5</i>	<i>PtALDH6B7</i>	O	0.189	1.627	0.116	89.41
6	<i>PtALDH6B3</i>	<i>PtALDH6B5</i>	T	0.162	1.441	0.113	79.15
7	<i>PtALDH7B4</i>	<i>PtALDH7B5</i>	W	0.020	0.226	0.088	12.40
10	<i>PtALDH10A8</i>	<i>PtALDH10A9</i>	W	0.034	0.153	0.221	8.38
11	<i>PtALDH11A3</i>	<i>PtALDH11A4</i>	W	0.018	0.188	0.096	10.35
18	<i>PtALDH18B1</i>	<i>PtALDH18B2</i>	W	0.042	0.270	0.154	14.85

Notes: Gene pairs were identified at the terminal nodes of the phylogenetic tree shown in Fig 1. Synonymous (Ks) and nonsynonymous substitution (Ka) rates are presented for each pair. Gene pairs created by tandem duplication (T), whole genome duplication (W), or other (O) events are indicated in the table.

doi:10.1371/journal.pone.0124669.t002

when a recent large-scale genome duplication event is thought to have occurred in *Populus* [53].

Evolutionary relationship between the *Populus* and *Arabidopsis* ALDH gene families

By comparing the genome sequences from different taxa it is possible to reconstruct the evolutionary history of each gene in its entirety [54]. To further investigate the origin and evolution of *Populus* ALDHs, we analyzed a comparative synteny map of the *Populus* and *Arabidopsis* genomes (Fig 3). *Arabidopsis* is an important model plant species and the functions of most *Arabidopsis* ALDH genes have been well characterized [8]. Thus, the comparative genomics analysis allowed us to infer the functions of the *Populus* ALDHs based on the annotations of their homologs in *Arabidopsis*.

Large-scale syntenies of orthologs from eight ALDH families (ALDH2, 3, 5, 7, 10, 11, 18, and 22) in the *Arabidopsis* and *Populus* genomes were identified (Fig 3). The syntenies were unambiguous and included the following orthologous pairs: *AtALDH2C4*-*PtALDH2C4*, *AtALDH3F1*-*PtALDH3F1*, *AtALDH5F1*-*PtALDH5F1* and *AtALDH22A1*-*PtALDH22A1* (Fig 3), indicating that these genes/families were in the genome of the last common ancestor of *Populus* and *Arabidopsis*. Three of the *Populus* ALDH paralogous gene pairs were syntenic with *Arabidopsis* ALDHs (*AtALDH3H1*-*PtALDH3H1/3H4*, *AtALDH7B4*-*PtALDH7B4/7B5* and *AtALDH11A3*-*PtALDH11A3/11A4*), and these three *PtALDH* gene pairs were probably duplicated in the recent large-scale genome duplication event (13 million years ago) in *Populus* (Table 2). The syntenic interpretation was more challenging where duplicated *Populus* genes corresponded to two *Arabidopsis* ALDHs (e.g. *AtALDH10A8/10A9*-*PtALDH10A8/10A9*, *AtALDH18B1/18B2*-*PtALDH18B1/18B2*). The remaining two families (ALDH6 and ALDH12) did not map to any of the synteny blocks. However, it was not possible to conclude that these two ALDH families from *Populus* and *Arabidopsis* did not share a common ancestor because, after speciation, the *Populus* and *Arabidopsis* genomes may have undergone multiple rounds of significant chromosomal rearrangement and fusions, followed by selective gene loss [23].

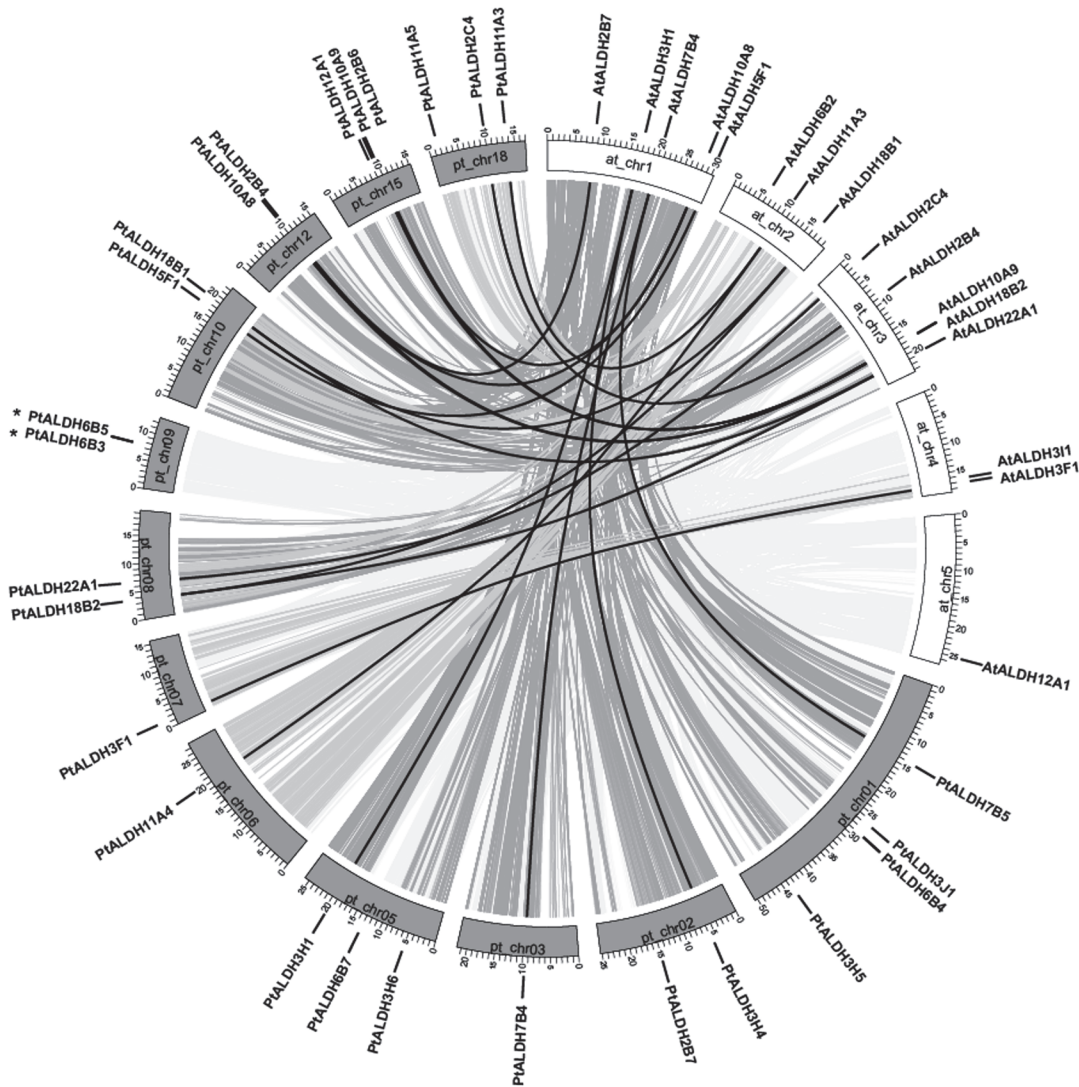


Fig 3. Synteny analysis of ALDH genes between *Populus* and *Arabidopsis*. *Populus* and *Arabidopsis* chromosomes are depicted as grey and blank boxes, respectively. ALDH genes are indicated by vertical black lines outer the circles. Grey lines connected between *Populus* and *Arabidopsis* chromosomes denoted syntenic relationships.

doi:10.1371/journal.pone.0124669.g003

Populus ALDH genes were differentially expressed in different tissues

Whole-genome microarray assays have been used successfully to study gene expression profiles in *Populus* [45, 47]. To gain insight into the expression patterns of *Populus* ALDH genes in different tissues, a comprehensive analysis was conducted based on an Affymetrix (GSE13990) and a Nimblegen (GSE13043) microarray data generated by Wilkins and Dharmawardhana [34, 35]. Although these two microarray datasets were generated on different platforms, they largely represent the *Populus* ALDHs presented in this study.

Most *Populus* ALDHs show distinct tissue-specific expression patterns. As shown in Fig 4A, five genes (*PtALDH3F1*, *3H5*, *3J1*, *11A3* and *11A4*) showed relatively high expression levels in young and mature leaves but low expression level in differentiating xylem, root, and male and female catkins. Four genes (*PtALDH6B4*, *6B5*, *10A8*, and *11A5*) were expressed mainly in root, three genes (*PtALDH3H1*, *18B1*, and *18B2*) had relatively high expression levels in male and

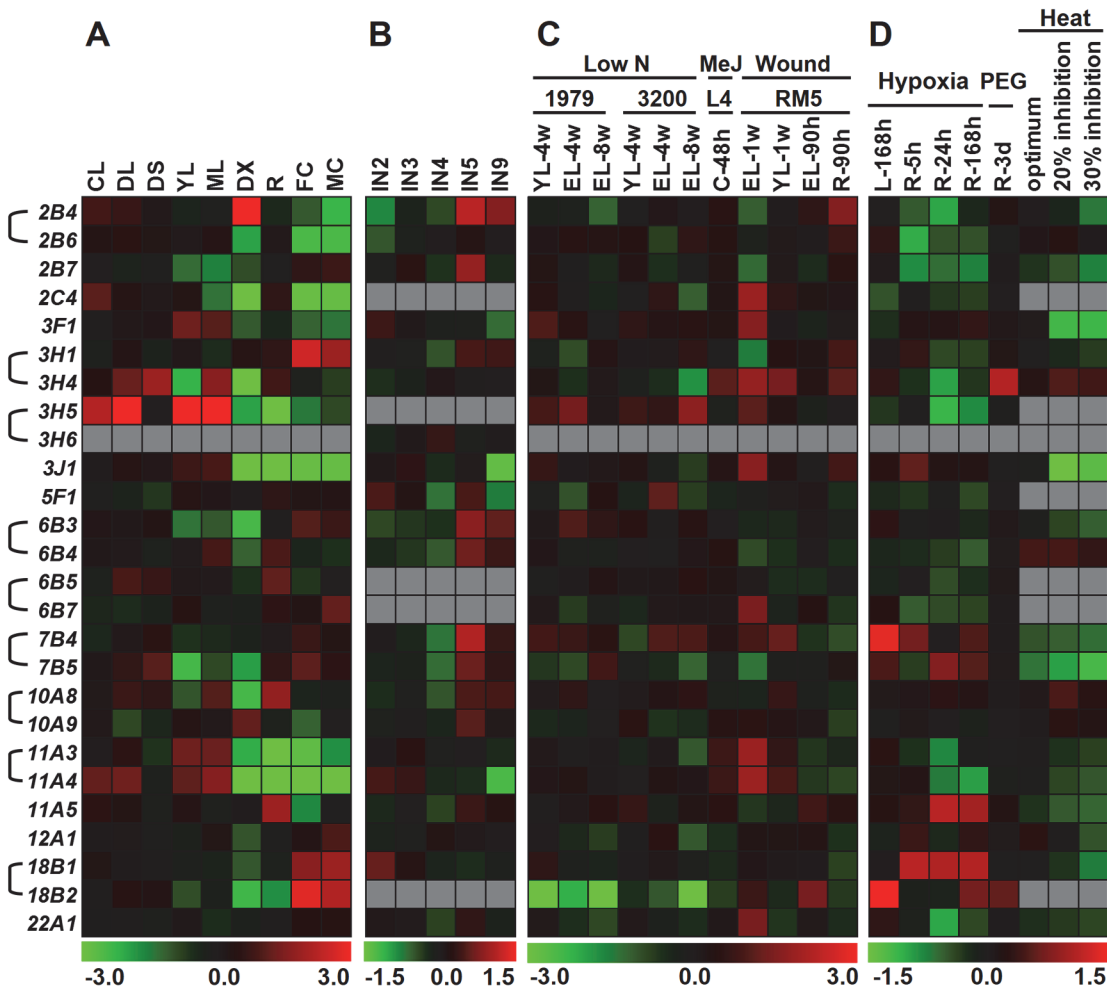


Fig 4. Expression profiles of *Populus* ALDH genes across different tissues and various abiotic stresses. (A) Heatmap showing expression profiles of *PtALDH* genes across various tissues analyzed. The Affymetrix microarray data were obtained from NCBI Gene Expression Omnibus (GEO) database under the series accession number GSE13990. CL, continuous light-grown seedling; DL, etiolated dark-grown seedling transferred to light for 3 h; DS, dark-grown seedlings; YL, young leaf; ML, mature leaf; DX, differentiating xylem; R, root; FC, female catkins; MC, male catkins. (B) Heatmap showing expression profiles of *PtALDH* genes at different stem development stages. The NimbleGen microarray data were obtained from NCBI GEO database under the series accession number GSE17230. IN2-IN9, stem internodes 2 to stem internodes 9. (C) Heatmap showing expression profiles of *PtALDH* genes across various stresses and genotypes analyzed. Microarray data under the series accession number GSE16786 was obtained from NCBI GEO database. Genotypes analyzed included: *P. fremontii* × *angustifolia* clones 1979, 3200, and RM5, *P. tremuloides* clones 271 and L4, and *P. deltoides* clones Soligo and Carpaccio. Tissues analyzed included: YL, young leaves; EL, expanding leaves; R, root tips; C, suspension cell cultures. Stress treatments included: low N, nitrogen limitation; MeJ, Methyl Jasmonate elicitation; Wound, sampled either one week or 90 hours after wounding. (D) Heatmap showing expression profiles of *PtALDH* genes under hypoxia, drought and heat stresses. Microarray data under the series accession number GSE13109 (hypoxia), GSE17225 (drought) and GSE26199 (heat) was obtained from NCBI GEO database. Stress treatments included: Hypoxia, the root system of grey poplar (*Populus* × *canescens*) were flooded for up to 168 h; Drought, a moderate water deficit was applied by adding PEG to the nutrient solution (200g/l) on hybrid poplar (*P. deltoides* × *P. nigra*) grown in hydroponics; Heat, fully expanded leaf samples of *P. trichocarpa* were harvested at 4 physiological states as determined from prior gas exchange measurements (growth temperature, 22°C—baseline, 31.75°C—photosynthetic optimum, 38.4°C—20% inhibition of optimum and 40.5°C—30% inhibition of optimum). Color scale represents log₂ expression values, green represents low level and red indicates high level of transcript abundances.

doi:10.1371/journal.pone.0124669.g004

female catkins, and only *PtALDH2B4* and *10A9* exhibited high expression levels in differentiating xylem. The tissue-specific expression patterns for these genes implied their involvement in special developmental processes. Genes in single-member families (*PtALDH5F1*, *12A1*, and *22A1*, Fig 4A) tended to maintain consistent expression levels across diverse organs, probably because of functional constraints, suggesting that these genes may participate in the basic metabolism of *Populus*.

To identify putative *Populus* ALDHs involved in stem development, a heat map was generated based on the microarray data (GSE13043). As shown in Fig 4B, most of the ALDHs exhibited different expression levels in *Populus* stem segments (IN2-IN5, and IN9), where IN2 and IN3 represent the vascular tissues of primary growth, mainly primary xylem and primary phloem, and IN5 and IN9 represent well developed secondary phloem and secondary xylem vessels, as well as fibers with well lignified secondary cell walls [35]. The expression profiles of these ALDHs suggested they may play special roles during each stage of cell wall biosynthesis. Of the *Populus* ALDHs we examined in this study, 10 were predominantly expressed in IN5 (*PtALDH2B4*, *2B7*, *3H1*, *6B3*, *6B4*, *7B4*, *7B5*, *10A8*, *10A9*, and *11A5*), suggesting they may be involved in the transition from primary to secondary growth (Fig 4B). The expression patterns of nine *PtALDH* paralogous gene pairs in five tissues (YL—young leaf, ML—mature leaf, PS—primary stem, SS—secondary stem, and R—root) were also examined by qRT-PCR and the mRNA levels of the detected genes were generally consistent with the results from the microarray data. Nair et al. [55] reported that *AtALDH2C4* (*REF1*, At3g24503) was involved in the formation of both soluble and cell wall-linked ferulate esters. In maize, the ortholog of *Arabidopsis* *ALDH22A1* was highly expressed in caffeic acid *O*-methyltransferase deficient tissues, and was also the most expressed ALDH in normal internodes [56]. Similar to *AtALDH2C4* in *Arabidopsis*, the ortholog *ZmALDH2C2* (*RF2C*) in maize also was reported to be involved in the biosynthesis of ferulic acid, a major esterified hydroxycinnamic acid in cell walls that impedes the hydrolysis of the cell wall biomass [57].

Populus ALDH genes potentially involved in response to abiotic stresses

Plant ALDHs have been reported to play important roles in the adaptation of plants to various abiotic stresses [8, 18]. Here, we analyzed the expression profiles of *PtALDHs* under abiotic stresses such as low nitrogen, methyl jasmonate (MeJ) treatment, mechanical wounding, hypoxia, drought, and heat (series accession numbers GSE13109, GSE17225, GSE26199 and GSE16786) [36, 37]. *PtALDH18B1* was commonly down-regulated under nitrogen deprivation stress in 4-week-old young leaves, 4-week and 8-week-old expanded leaves in two different *Populus* genotypes (1979 and 3200, Fig 4C). Other *PtALDHs* showed different response profiles to nitrogen deficit stress between these two *Populus* genotypes. For instance, *PtALDH3H5* and *7B4* were significantly up-regulated in 8-week-old expanded leaves in genotype 3200, whereas no distinctive expression patterns were observed in genotype 1979. In response to MeJ feeding in cell culture, three genes (*PtALDH3H4*, *11A3*, and *11A4*) were found to be up-regulated (Fig 4C). Mechanical wounding stress commonly caused up-regulation of 12 genes at 1 week after wounding in expanded leaves. In addition, five genes (*PtALDH2B4*, *2B6*, *3J1*, *3H1*, and *3H4*) were up-regulated at 90 hours after wounding in root tips (R), suggesting the functional divergence of *PtALDHs* in response to mechanical wounding.

In response to hypoxia, two genes (*PtALDH7B4* and *18B2*) were up-regulated significantly in leaves at 168 hours after hypoxia, while three other genes (*PtALDH7B5*, *11A5*, and *18B1*) were up-regulated during hypoxia in roots (Fig 4D). Drought stress caused up-regulation of two genes (*PtALDH3H4* and *18B2*) at 3 days after polyethylene glycol (PEG) in roots (Fig 4D). In a previous study, the physiological condition was divided into four states based on the *Populus* photosynthetic activity at temperatures from 22°C to 42°C: baseline (22°C, the growth temperature), optimum (31.75°C, temperature at which the maximum net CO₂ assimilation rate is observed), 20% inhibition of optimum (38.4°C), and 30% inhibition of optimum (40.5°C) [37]. Most *PtALDHs* were down-regulated under heat stress and only two genes (*PtALDH3H4* and *6B4*) were up-regulated when photosynthesis was inhibited by 20% and 30% (Fig 4D). ALDHs that were induced under various stresses have been identified in many plant species, indicating

that they may play critical roles in plant adaptation to these stresses [58]. In *Arabidopsis*, it was reported that overexpression of *ALDH3I1* may improve the plant's tolerance to diverse stresses [13]. In addition, both *ALDH3* and *ALDH7* were found to be involved in stress-regulated detoxification pathways. In *Arabidopsis*, the chloroplastic *ALDH3I1* and the cytoplasmic *ALDH7B4* may confer tolerance to osmotic and oxidative stresses [14]. *ALDH18* gene encodes P5CS (Δ^1 -pyrroline-5-carboxylate synthetase), a key regulatory enzyme that plays a crucial role in proline biosynthesis. Recent studies indicated that *P5CS1* was required for proline biosynthesis under osmotic stress in *Arabidopsis*, suggesting *ALDH18* may also be abiotic stress responsive [59]. Our findings are largely consistent with studies in *Arabidopsis*, rice, and grape that indicated *ALDH* genes from families 2, 3, 6, 7, 11, and 18 were significantly induced in abiotic stressed plants [10, 14, 18, 58, 59]. Our expression data indicated that some *Populus* *ALDH* genes are potential candidates for improving *Populus* tolerance to abiotic stresses. Extensive further studies are also required to examine the exact biochemical roles of *Populus* *ALDHs* in developmental processes and stress tolerance.

Divergent expression of *PtALDH* gene pairs

Previous studies of some closely-related *ALDHs* hinted at potential roles of functional specialization in the retention of duplicated genes [11, 60]. Most *Populus* *ALDHs* arose from recent genome duplication and tandem duplication events (Table 2). Duplicated *ALDHs* showed different tissue-specific expression patterns (Fig 4), suggesting that gene duplications supplied opportunities for the duplicates to be free from the functional constraints of the parent gene. To analyze the expression divergence between *PtALDH* paralogous gene pairs, we identified the putative *cis*-acting elements in the promoter regions of nine *PtALDH* gene pairs (from -1,000 nt upstream to +200 nt downstream of the transcription start site) using PlantCARE database (Fig 5). The nine *PtALDH* gene pairs all had different *cis*-acting elements in their promoter regions. We also examined the mRNA levels of the *PtALDH* gene pairs in different tissues by qRT-PCR to validate their expression patterns (Fig 6). For the *PtALDH2B4/PtALDH2B6* pair, several leaf development-related *cis*-acting elements (three as-2-boxes: involved in shoot-specific expression and light responsiveness, one HD-Zip1: involved in differentiation of the palis and mesophyll cells, and one HD-Zip2: involved in the control of leaf morphology development) were detected in the promoter of *PtALDH2B6* (Fig 5) but not in the *PtALDH2B4* promoter. As expected based on this finding, *PtALDH2B6* was highly expressed in young and mature leaves (Fig 6A), whereas *PtALDH2B4* had low mRNA levels because it lacked leaf development-related *cis*-acting element in its promoter. For the *PtALDH3H1/PtALDH3H4* pair, six HSE (involved in heat stress responsiveness) and one MBS (MYB binding site involved in drought-inducibility) were detected in the promoter of *PtALDH3H4* (Fig 5) but not in the promoter of *PtALDH3H1*; therefore, *PtALDH3H4* was significantly induced by heat and drought (PEG treatment) stresses, whereas *PtALDH3H1* was not (Fig 4D). In addition, many hormone related *cis*-acting elements, development related *cis*-acting elements, and stress response *cis*-acting elements were detected in the promoters of various *PtALDHs*, implying that different members of the *Populus* *ALDH* families were involved in different development processes and stress responses. Based on these data, we propose that expression divergence and/or functional specialization may have played important roles in the retention of the *Populus* *ALDH* duplicate genes. However, the functions of the *PtALDH* genes associated with development and stress responses in *Populus* need to be investigated further.

Functional diversification among gene family members is considered an important source of evolutionary innovation in complex organisms, and various theoretical models have been proposed to explain the mechanisms involved [61–64]. The most plausible models proposed

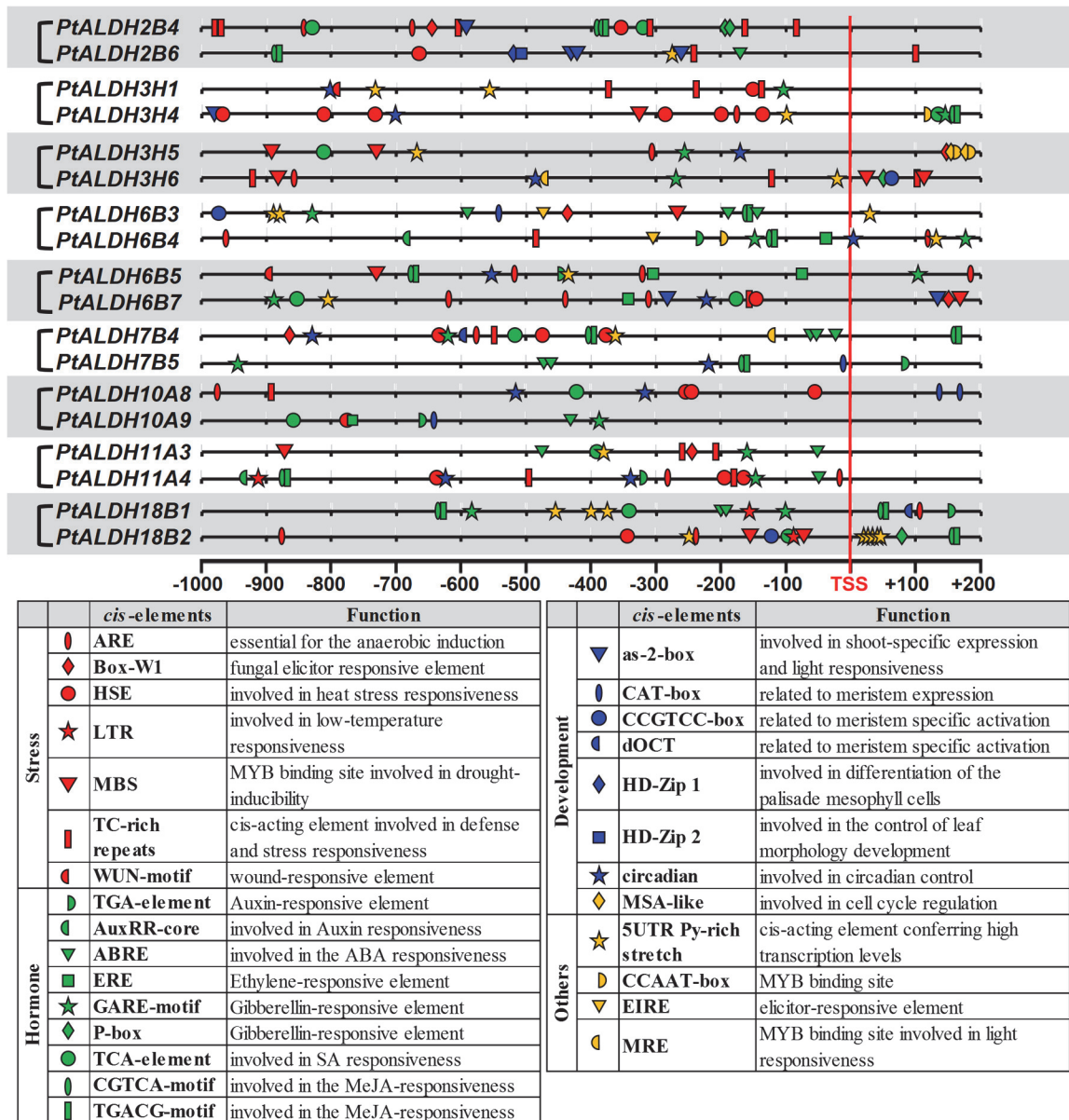


Fig 5. Comparisons of the *cis*-acting elements in the promoter regions of *PtALDH* gene pairs. Schematic representation of nine *PtALDH* gene pairs promoters. A selection of putative *cis*-acting elements identified in the PlantCARE database is indicated. Distances shown in the figure are relative to the Transcription Start Site (TSS).

doi:10.1371/journal.pone.0124669.g005

for the retention of duplicated genes invoke sub- or neo-functionalization [24]. In this study, clear divergence in expression patterns was observed among the *Populus* *ALDHs* in different tissues and in response to different stress treatments. These findings clearly support the assertion that expression divergence is often the first step in functional divergence between duplicate genes and that this divergence increases the chance of duplicate genes being retained in a genome [61]. Our findings provide evidence for the evolutionary partitioning of ancestral functions among duplicated genes.

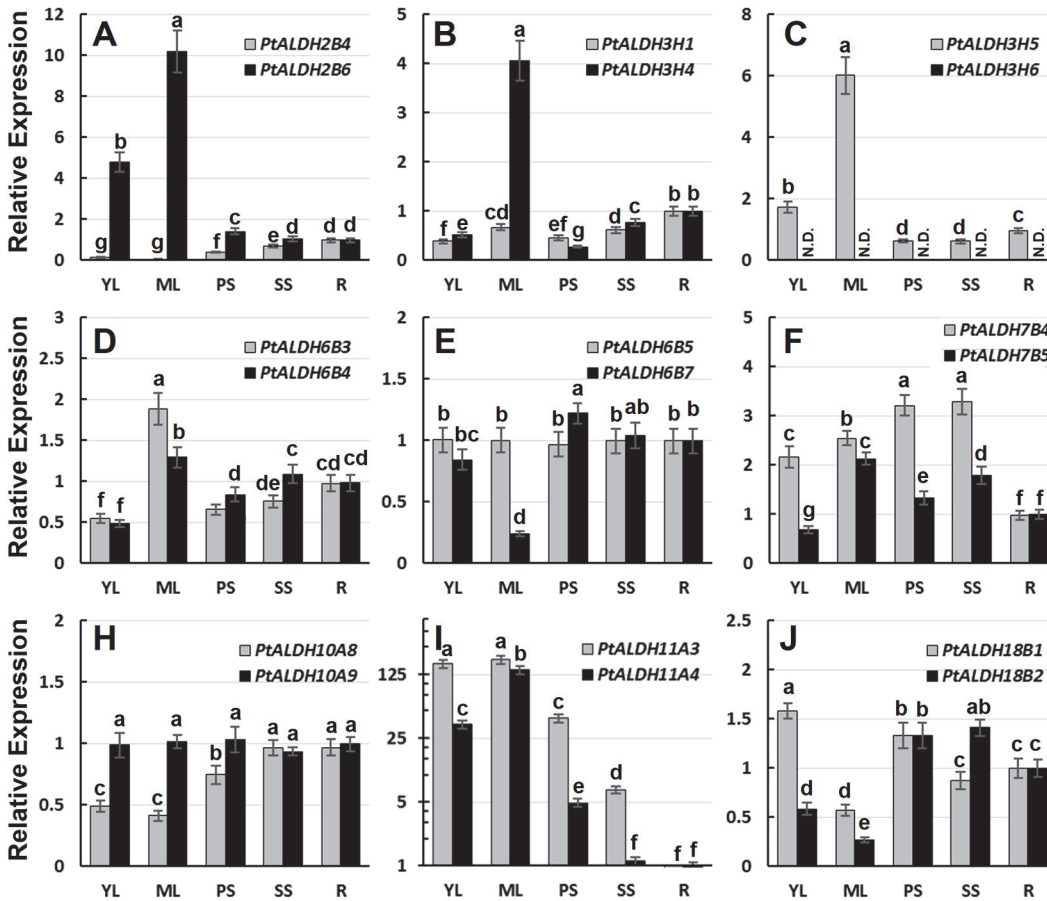


Fig 6. Expression analysis of nine *PtALDH* gene pairs in different tissues using qRT-PCR. The relative mRNA abundance of nine *PtALDH* gene pairs were quantified in vegetative tissues (YL, young leaves; ML, mature leaves; PS, primary stem; SS, secondary stem; R, roots). The average expression of each gene was calculated relatively to the first biological replicate of roots \pm standard error (SE) ($n \geq 3$). Relative expression represents \log_2 expression values. Bars with the same letters are not significantly different according to Duncan test and Fisher's protected LSD test ($P < 0.05$).

doi:10.1371/journal.pone.0124669.g006

Conclusions

Aldehyde dehydrogenases (*ALDH*s) are members of the NAD(P)^+ -dependent protein superfamily that catalyze the oxidation of a wide range of endogenous and exogenous highly reactive aliphatic and aromatic aldehyde molecules. Although the *ALDH* gene superfamily has been reviewed in many plants, no systematic analyses have been conducted to date in *Populus*, a model tree. In the present study, comprehensive analyses including phylogeny, gene structure, chromosomal location, expression profiles, and the *cis*-acting elements of members of the *Populus ALDH* gene superfamily were performed. A total of 26 *Populus ALDH* genes were grouped into 10 families. We found that the exon-intron structures were relatively conserved within each family. Comparative analysis showed that 10 paralogous gene pairs were created by different duplication types. An additional comprehensive analysis of the expression profiles provided insights into the possible functional divergence among members of the *ALDH* gene superfamily. Gene specific promoter *cis*-acting elements may explain the divergent expression patterns observed between eight of nine *PtALDH* gene pairs (one pair may have been generated by tandem duplication). Although the functions of the *PtALDH*s remain largely unknown and many experiments are needed to determine their exact functions, our phylogenetic and

expression analyses may help in the selection of appropriate candidate genes for further functional characterization.

Supporting Information

S1 Fig. The specificity of probe sets between gene pairs.

(PDF)

S1 Table. List of ALDH protein sequences identified from *P. trichocarpa* and *A. thaliana*.

(XLSX)

S2 Table. Probe sets corresponding to *Populus* ALDH genes.

(XLSX)

S3 Table. The sequences of qRT-PCR primers.

(DOCX)

S4 Table. Amino acid identity among *P. trichocarpa* and *A. thaliana* ALDH proteins was analyzed in pairwise fashion.

(XLSX)

S5 Table. Number of ALDH family members identified in various organisms.

(DOCX)

Acknowledgments

This work was supported by National Nonprofit Institute Research Grant of CAF [CAFYBB2014ZX001-4] and [TGB2013009] to JJH and the China Postdoctoral Science Foundation [2014M550104] to JZ. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author Contributions

Conceived and designed the experiments: JZ JJH. Performed the experiments: FXT TW JZ. Analyzed the data: FXT JLZ YLX JZ JJH. Contributed reagents/materials/analysis tools: FXT JZ JJH. Wrote the paper: FXT JZ JJH.

References

1. Vasiliou V, Pappa A, Petersen DR. Role of aldehyde dehydrogenases in endogenous and xenobiotic metabolism. *Chemico-Biol Interac.* 2000; 129(1):1–19.
2. Yoshida A, Rzhetsky A, Hsu LC, Chang C. Human aldehyde dehydrogenase gene family. *Eur J Biochem.* 1998; 251(3):549–57. PMID: [9490025](#)
3. Lindahl R. Aldehyde dehydrogenases and their role in carcinogenesis. *Crit Rev Biochem Mol.* 1992; 27(4–5):283–335.
4. Bartels D. Targeting detoxification pathways: an efficient approach to obtain plants with multiple stress tolerance? *Trends Plant Sci.* 2001; 6(7):284–6. PMID: [11435150](#)
5. Kotchoni S, Bartels D. Water stress induces the up-regulation of a specific set of genes in plants: aldehyde dehydrogenase as an example. *Bulg J Plant Physiol Special.* 2003;(2003):37–51.
6. Vasiliou V, Bairoch A, Tipton KF, Nebert DW. Eukaryotic aldehyde dehydrogenase (ALDH) genes: human polymorphisms, and recommended nomenclature based on divergent evolution and chromosomal mapping. *Pharmacogenet Genom.* 1999; 9(4):421–34.
7. Sophos NA, Vasiliou V. Aldehyde dehydrogenase gene superfamily: the 2002 update. *Chemico-Biol Interac.* 2003; 143:5–22.
8. Kirch H-H, Bartels D, Wei Y, Schnable PS, Wood AJ. The ALDH gene superfamily of *Arabidopsis*. *Trends Plant Sci.* 2004; 9(8):371–7. PMID: [15358267](#)

9. Skibbe DS, Liu F, Wen T-J, Yandea MD, Cui X, Cao J, et al. Characterization of the aldehyde dehydrogenase gene families of *Zea mays* and *Arabidopsis*. *Plant Mol Biol*. 2002; 48(5–6):751–64. PMID: [11999847](#)
10. Zhang Y, Mao L, Wang H, Brocker C, Yin X, Vasiliou V, et al. Genome-wide identification and analysis of grape aldehyde dehydrogenase (*ALDH*) gene superfamily. *PLoS One*. 2012; 7(2):e32153. doi: [10.1371/journal.pone.0032153](#) PMID: [22355416](#)
11. Liu F, Schnable PS. Functional specialization of maize mitochondrial aldehyde dehydrogenases. *Plant Physiol*. 2002; 130(4):1657–74. PMID: [12481049](#)
12. Shin J-H, Kim S-R, An G. Rice aldehyde dehydrogenase7 is needed for seed maturation and viability. *Plant Physiol*. 2009; 149(2):905–15. doi: [10.1104/pp.108.130716](#) PMID: [19052152](#)
13. Sunkar R, Bartels D, Kirch HH. Overexpression of a stress-inducible aldehyde dehydrogenase gene from *Arabidopsis thaliana* in transgenic plants improves stress tolerance. *Plant J*. 2003; 35(4):452–64. PMID: [12904208](#)
14. Kotchoni SO, Kuhns C, Ditzer A, KIRCH HH, Bartels D. Over-expression of different aldehyde dehydrogenase genes in *Arabidopsis thaliana* confers tolerance to abiotic stress and protects plants against lipid peroxidation and oxidative stress. *Plant Cell Environ*. 2006; 29(6):1033–48. PMID: [17080931](#)
15. Rodrigues SM, Andrade MO, Gomes APS, DaMatta FM, Baracat-Pereira MC, Fontes EP. *Arabidopsis* and tobacco plants ectopically expressing the soybean antiquitin-like ALDH7 gene display enhanced tolerance to drought, salinity, and oxidative stress. *J Exp Bot*. 2006; 57(9):1909–18. PMID: [16595581](#)
16. Li X, Guo R, Li J, Singer SD, Zhang Y, Yin X, et al. Genome-wide identification and analysis of the aldehyde dehydrogenase (*ALDH*) gene superfamily in apple (*Malus domestica* Borkh.). *Plant Physiol Biochem*. 2013; 71:268–82. doi: [10.1016/j.plaphy.2013.07.017](#) PMID: [23978559](#)
17. Huang W, Ma X, Wang Q, Gao Y, Xue Y, Niu X, et al. Significant improvement of stress tolerance in tobacco plants by overexpressing a stress-responsive aldehyde dehydrogenase gene from maize (*Zea mays*). *Plant Mol Biol*. 2008; 68(4–5):451–63. doi: [10.1007/s11103-008-9398-1](#) PMID: [18830798](#)
18. Gao C, Han B. Evolutionary and expression study of the aldehyde dehydrogenase (*ALDH*) gene superfamily in rice (*Oryza sativa*). *Gene*. 2009; 431(1):86–94.
19. Hofberger JA, Nsibo DL, Govers F, Bouwmeester K, Schranz ME. A Complex Interplay of Tandem-and Whole Genome Duplication Drives Expansion of the L-type Lectin Receptor Kinase Gene Family in the Brassicaceae. *Genome Biol Evol*. 2015:evv020.
20. Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, et al. Ancestral polyploidy in seed plants and angiosperms. *Nature*. 2011; 473(7345):97–100. doi: [10.1038/nature09916](#) PMID: [21478875](#)
21. Fawcett JA, Maere S, Van de Peer Y. Plants with double genomes might have had a better chance to survive the Cretaceous–Tertiary extinction event. *P Natl Acad Sci USA*. 2009; 106(14):5737–42. doi: [10.1073/pnas.0900906106](#) PMID: [19325131](#)
22. Roth C, Rastogi S, Arvestad L, Dittmar K, Light S, Ekman D, et al. Evolution after gene duplication: models, mechanisms, sequences, systems, and organisms. *J Exp Zool Part B*. 2007; 308(1):58–73. PMID: [16838295](#)
23. Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, et al. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science*. 2006; 313(5793):1596–604. PMID: [16973872](#)
24. Lan T, Yang Z-L, Yang X, Liu Y-J, Wang X-R, Zeng Q-Y. Extensive functional diversification of the *Populus* glutathione S-transferase supergene family. *Plant Cell*. 2009; 21(12):3749–66. doi: [10.1105/tpc.109.070219](#) PMID: [19996377](#)
25. Brocker C, Vasiliou M, Carpenter S, Carpenter C, Zhang Y, Wang X, et al. Aldehyde dehydrogenase (*ALDH*) superfamily in plants: gene nomenclature and comparative genomics. *Planta*. 2013; 237(1):189–210. doi: [10.1007/s00425-012-1749-0](#) PMID: [23007552](#)
26. Horton P, Park KJ, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, et al. WoLF PSORT: protein localization predictor. *Nucleic Acids Res*. 2007; 35(Web Server issue):W585–7. PMID: [17517783](#)
27. Larkin M, Blackshields G, Brown N, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. *Bioinformatics*. 2007; 23(21):2947–8. PMID: [17846036](#)
28. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol*. 1987; 4(4):406–25. PMID: [3447015](#)
29. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol*. 2011; 28(10):2731–9. doi: [10.1093/molbev/msr121](#) PMID: [21546353](#)
30. Guo AY, Zhu QH, Chen X, Luo JC. GSDBS: a gene structure display server. *Yi Chuan*. 2007; 29(8):1023–6. PMID: [17681935](#)

31. Lee T-H, Tang H, Wang X, Paterson AH. PGDD: a database of gene and genome duplication in plants. *Nucleic Acids Res.* 2013; 41(D1):D1152–D8. doi: [10.1093/nar/gks1104](https://doi.org/10.1093/nar/gks1104) PMID: [23180799](https://pubmed.ncbi.nlm.nih.gov/23180799/)
32. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: an information aesthetic for comparative genomics. *Genome Res.* 2009; 19(9):1639–45. doi: [10.1101/gr.092759.109](https://doi.org/10.1101/gr.092759.109) PMID: [19541911](https://pubmed.ncbi.nlm.nih.gov/19541911/)
33. Barrett T, Edgar R. Gene Expression Omnibus: Microarray Data Storage, Submission, Retrieval, and Analysis. *Method Enzymol.* 2006; 411:352–69. PMID: [16939800](https://pubmed.ncbi.nlm.nih.gov/16939800/)
34. Wilkins O, Nahal H, Foong J, Provart NJ, Campbell MM. Expansion and diversification of the *Populus* R2R3-MYB family of transcription factors. *Plant Physiol.* 2009; 149(2):981–93. doi: [10.1104/pp.108.132795](https://doi.org/10.1104/pp.108.132795) PMID: [19091872](https://pubmed.ncbi.nlm.nih.gov/19091872/)
35. Dharmawardhana P, Brunner AM, Strauss SH. Genome-wide transcriptome analysis of the transition from primary to secondary stem development in *Populus trichocarpa*. *BMC Genomics.* 2010; 11(1):150.
36. Kreuzwieser J, Hauberg J, Howell KA, Carroll A, Rennenberg H, Millar AH, et al. Differential response of gray poplar leaves and roots underpins stress adaptation during hypoxia. *Plant Physiol.* 2009; 149(1):461–73. doi: [10.1104/pp.108.125989](https://doi.org/10.1104/pp.108.125989) PMID: [19005089](https://pubmed.ncbi.nlm.nih.gov/19005089/)
37. Weston DJ, Karve AA, Gunter LE, Jawdy SS, Yang X, Allen SM, et al. Comparative physiology and transcriptional networks underlying the heat shock response in *Populus trichocarpa*, *Arabidopsis thaliana* and *Glycine max*. *Plant Cell Environ.* 2011; 34(9):1488–506. doi: [10.1111/j.1365-3040.2011.02347.x](https://doi.org/10.1111/j.1365-3040.2011.02347.x) PMID: [21554326](https://pubmed.ncbi.nlm.nih.gov/21554326/)
38. Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP. Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res.* 2003; 31(4):e15–e. PMID: [12582260](https://pubmed.ncbi.nlm.nih.gov/12582260/)
39. Jimenez-Lopez JC, Gachomo EW, Seufferheld MJ, Kotchoni SO. The maize ALDH protein superfamily: linking structural features to functional specificities. *BMC Struct Biol.* 2010; 10(1):43.
40. Wood AJ, Duff RJ. The aldehyde dehydrogenase (*ALDH*) gene superfamily of the moss *Physcomitrella patens* and the algae *Chlamydomonas reinhardtii* and *Ostreococcus tauri*. *Bryologist.* 2009; 112(1):1–11.
41. García-Ríos M, Fujita T, LaRosa PC, Locy RD, Clithero JM, Bressan RA, et al. Cloning of a polycistronic cDNA from tomato encoding γ -glutamyl kinase and γ -glutamyl phosphate reductase. *P Natl Acad Sci USA.* 1997; 94(15):8249–54. PMID: [9223347](https://pubmed.ncbi.nlm.nih.gov/9223347/)
42. Zhou M-L, Zhang Q, Zhou M, Qi L-P, Yang X-B, Zhang K-X, et al. Aldehyde dehydrogenase protein superfamily in maize. *Funct Integr Genomic.* 2012; 12(4):683–91. doi: [10.1007/s10142-012-0290-3](https://doi.org/10.1007/s10142-012-0290-3) PMID: [22983498](https://pubmed.ncbi.nlm.nih.gov/22983498/)
43. Hardison RC. A brief history of hemoglobins: plant, animal, protist, and bacteria. *P Natl Acad Sci USA.* 1996; 93(12):5675. PMID: [8650150](https://pubmed.ncbi.nlm.nih.gov/8650150/)
44. Brunner AM, Busov VB, Strauss SH. Poplar genome sequence: functional genomics in an ecologically dominant plant species. *Trends Plant Sci.* 2004; 9(1):49–56. PMID: [14729219](https://pubmed.ncbi.nlm.nih.gov/14729219/)
45. Chai G, Hu R, Zhang D, Qi G, Zuo R, Cao Y, et al. Comprehensive analysis of CCH zinc finger family in poplar (*Populus trichocarpa*). *BMC Genomics.* 2012; 13(1):253. doi: [10.1186/1471-2164-13-253](https://doi.org/10.1186/1471-2164-13-253)
46. Hu R, Qi G, Kong Y, Kong D, Gao Q, Zhou G. Comprehensive analysis of NAC domain transcription factor gene family in *Populus trichocarpa*. *BMC Plant Biol.* 2010; 10:145. doi: [10.1186/1471-2229-10-145](https://doi.org/10.1186/1471-2229-10-145) PMID: [20630103](https://pubmed.ncbi.nlm.nih.gov/20630103/)
47. Zhang J, Li J, Liu B, Zhang L, Chen J, Lu M. Genome-wide analysis of the *Populus Hsp90* gene family reveals differential expression patterns, localization, and heat stress responses. *BMC Genomics.* 2013; 14(1):1–14.
48. Prince VE, Pickett FB. Splitting pairs: the diverging fates of duplicated genes. *Nat Rev Genet.* 2002; 3(11):827–37. PMID: [12415313](https://pubmed.ncbi.nlm.nih.gov/12415313/)
49. Vandepoele K, Simillion C, Van de Peer Y. Evidence that rice and other cereals are ancient aneuploids. *Plant Cell.* 2003; 15(9):2192–202. PMID: [12953120](https://pubmed.ncbi.nlm.nih.gov/12953120/)
50. Hurst LD. The *Ka/Ks* ratio: diagnosing the form of sequence evolution. *Trends Genet.* 2002; 18(9):486–7. PMID: [12175810](https://pubmed.ncbi.nlm.nih.gov/12175810/)
51. Nuruzzaman M, Sharoni AM, Satoh K, Al-Shammari T, Shimizu T, Sasaya T, et al. The thioredoxin gene family in rice: Genome-wide identification and expression profiling under different biotic and abiotic treatments. *Biochem Bioph Res Co.* 2012; 423(2):417–23.
52. Lynch M, Conery JS. The evolutionary fate and consequences of duplicate genes. *Science.* 2000; 290(5494):1151–5. PMID: [11073452](https://pubmed.ncbi.nlm.nih.gov/11073452/)
53. Sterck L, Rombauts S, Jansson S, Sterky F, Rouzé P, Van de Peer Y. EST data suggest that poplar is an ancient polyploid. *New Phytol.* 2005; 167(1):165–70. PMID: [15948839](https://pubmed.ncbi.nlm.nih.gov/15948839/)

54. Koonin EV. Orthologs, paralogs, and evolutionary genomics. *Annu Rev Genet.* 2005; 39:309–38. PMID: [16285863](#)
55. Nair RB, Bastress KL, Ruegger MO, Denault JW, Chapple C. The *Arabidopsis thaliana* REDUCED EPIDERMAL FLUORESCENCE1 gene encodes an aldehyde dehydrogenase involved in ferulic acid and sinapic acid biosynthesis. *Plant Cell.* 2004; 16(2):544–54. PMID: [14729911](#)
56. Guillaumie S, Goffner D, Barbier O, Martinant J-P, Pichon M, Barrière Y. Expression of cell wall related genes in basal and ear internodes of silking brown-midrib-3, caffeic acid O-methyltransferase (COMT) down-regulated, and normal maize plants. *BMC Plant Biol.* 2008; 8(1):71.
57. Bosch M, Mayer C-D, Cookson A, Donnison IS. Identification of genes involved in cell wall biogenesis in grasses by differential gene expression profiling of elongating and non-elongating maize internodes. *J Exp Bot.* 2011; 62(10):3545–61. doi: [10.1093/jxb/err045](#) PMID: [21402660](#)
58. Kirch H-H, Schlingensiepen S, Kotchoni S, Sunkar R, Bartels D. Detailed expression analysis of selected genes of the aldehyde dehydrogenase (*ALDH*) gene superfamily in *Arabidopsis thaliana*. *Plant Mol Biol.* 2005; 57(3):315–32. PMID: [15830124](#)
59. Székely G, Ábrahám E, Cséplő Á, Rigó G, Zsigmond L, Csiszár J, et al. Duplicated P5CS genes of *Arabidopsis* play distinct roles in stress regulation and developmental control of proline biosynthesis. *Plant J.* 2008; 53(1):11–28. PMID: [17971042](#)
60. Tsuji H, Tsutsumi N, Sasaki T, Hirai A, Nakazono M. Organ-specific expressions and chromosomal locations of two mitochondrial aldehyde dehydrogenase genes from rice (*Oryza sativa* L.), *ALDH2a* and *ALDH2b*. *Gene.* 2003; 305(2):195–204. PMID: [12609740](#)
61. Ohno S. *Evolution by gene duplication*: London: George Alien & Unwin Ltd. Berlin, Heidelberg and New York: Springer-Verlag.; 1970.
62. Moore RC, Purugganan MD. The evolutionary dynamics of plant duplicate genes. *Curr Opin Plant Biol.* 2005; 8(2):122–8. PMID: [15752990](#)
63. Hughes AL. The evolution of functionally novel proteins after gene duplication. *Proc R Soc Lond B Biol Sci.* 1994; 256(1346):119–24. PMID: [8029240](#)
64. Walsh B. Population-genetic models of the fates of duplicate genes. *Genetica.* 2003; 118(2–3):279–94. PMID: [12868606](#)